

XIV Международная научно-практическая конференция студентов аспирантов и молодых учёных
«Молодёжь и современные информационные технологии»

СТАТИСТИЧЕСКАЯ МОДЕЛЬ КЛАСТЕРИЗАЦИИ РОССИЙСКИХ ВУЗОВ В ФАКТОРНОМ ПРОСТРАНСТВЕ ВСТУПИТЕЛЬНЫХ ИСПЫТАНИЙ ПО НАПРАВЛЕНИЮ ПОДГОТОВКИ ЯДЕРНЫЕ ФИЗИКА И ТЕХНОЛОГИИ

Новосельцева Д.А.

Научный руководитель: к.ф.-м.н., доцент Михальчук А.А.

Томский политехнический университет

e-mail: dary_2503@mail.ru

Введение

В настоящее время в связи с введением федеральных государственных образовательных стандартов (ФГОС) высшего профессионального образования (ВПО) третьего поколения весьма актуальным является вопрос о компьютерном анализе вступительных испытаний абитуриентов. Анализ содержания ФГОС ВПО позволяет сделать вывод, что приоритетной задачей управления качеством образования на современном этапе выступает оценка качества освоения основных общеобразовательных программ, включающая результаты единого государственного экзамена (ЕГЭ). Контроль качества общего образования, в частности результатов ЕГЭ, даёт возможность оценить качество набора абитуриентов, приобретает характер мониторинга или непрерывного контроля для постоянного отслеживания результатов образования, хода образовательного процесса, необходимого для систематической корректировки мероприятий по их реализации. Целью данной работы является кластеризация российских вузов в факторном пространстве показателей вступительных испытаний (ПВИ) по направлению подготовки ядерные физика и технологии (НПФ).

Результаты статистического анализа

В качестве базы данных взяты данные 16-ти ВУЗов по следующим 8-ми показателям: $ЕГЭ_{15}$ - средний балл ЕГЭ зачисленных в 2015 году, $ЕГЭ_{14}$ - средний балл ЕГЭ зачисленных в 2014 году, $ДЕГЭ$ - динамический показатель (разность $ЕГЭ$ 2015 и 2014 гг.), N - количество студентов, зачисленных на бюджетные места, N_C - количество студентов, зачисленных на бюджетные места по результатам конкурса, N_{TS} - количество студентов, зачисленных на целевые места, N_L - количество студентов, зачисленных на бюджетные места по льготам, N_O - количество студентов, зачисленных на бюджетные места по результатам олимпиады. Таким образом, созданная в MS Excel база данных впоследствии использовалась в пакете Statistica [2] для статистического анализа данных.

Согласно проверке ПВИ на корреляционную зависимость были выявлены значимые корреляционные зависимости, на основе которых можно выделить 3 группы ПВИ $\{ДЕГЭ\}$, $\{ЕГЭ_{14}, ЕГЭ_{15}\}$, и $\{N, N_C, N_{TS}, N_L, N_O\}$. Причем различия между коэффициентами Спирмена R и Пирсона r

незначимые (например, для $ЕГЭ_{14}-ЕГЭ_{15}$ $r=0,90$ $R=0,93$ $p \approx 0,6 > 0,1$).

На рис. 1 изображена дендрограмма (древовидная диаграмма или иерархическое дерево) корреляционной матрицы ПВИ, полученная в результате древовидной кластеризации. В качестве меры близости ПВИ использовано корреляционное расстояние $1 - r$, где r - коэффициент корреляции Пирсона, а в качестве правила объединения для двух кластеров использован метод Уорда (Варда).

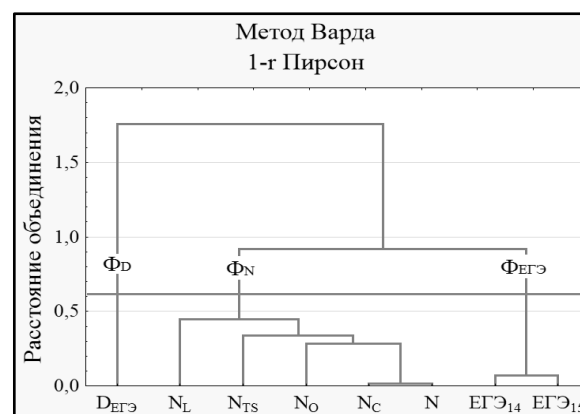


Рис.1. Горизонтальная дендрограмма корреляционной матрицы ПВИ

Наличие корреляционной связи ПВИ позволяет использовать факторный анализ для сокращения числа показателей и определения структуры взаимосвязей между ПВИ. С помощью метода главных компонент получена 3-мерная факторная модель. Факторные нагрузки первоначальных ПВИ разделились по факторам следующим образом: фактор 1 – фактор численности Φ_N характеризуется положительной корреляционной связью с N, N_C, N_L, N_{TS}, N_O и интерпретируется как фактор численности поступивших по НПФ; фактор 2 – фактор Φ_D характеризуется положительной корреляционной связью с $ДЕГЭ$ и интерпретируется как фактор динамики; фактор 3 – фактор вступительных испытаний $\Phi_{ЕГЭ}$ характеризуется положительной корреляционной связью с $ЕГЭ_{15}$ и $ЕГЭ_{14}$, и интерпретируется как фактор проходного балла по ЕГЭ (табл.1).

При проведении кластеризации вузов по НПФ методом древовидной кластеризации (рис.2) в построенном трехфакторном пространстве $\{\Phi_N, \Phi_D, \Phi_{ЕГЭ}\}$ в качестве мер близости выбраны Евклидово расстояние и расстояние городских

кварталов, а в качестве правила объединения двух кластеров использован метод Уорда. В результате кластеризации было получено устойчивое разбиение 16 ВУЗов на 7 кластеров.

Таблица 1. Матрица факторной структуры
ПВИ вузов по НПФ 2015 г.

Показатель	Φ_N	Φ_D	$\Phi_{ЕГЭ}$
$ЕГЭ_{15}$	0.280	0.06	0.949
$ЕГЭ_{14}$	0.263	-0.308	0.904
$D_{ЕГЭ}$	-0.020	0.984	-0.106
N	0.950	-0.041	0.261
N_C	0.929	-0.011	0.185
N_{TS}	0.772	-0.212	0.345
N_L	0.586	-0.222	0.554
N_O	0.779	0.119	0.448
Доля фактора	0.433	0.147	0.307

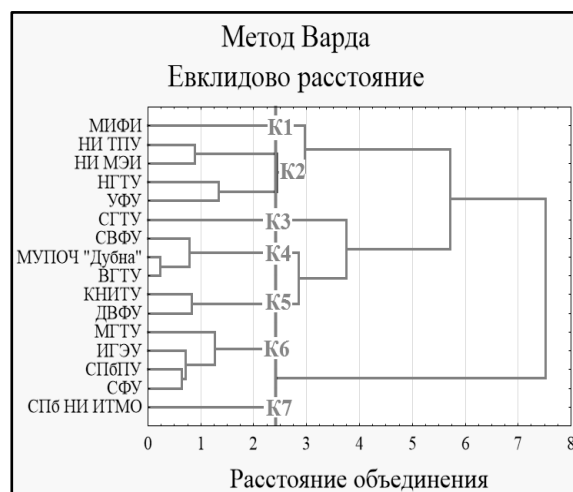


Рис.2. Горизонтальная дендрограмма вузов по
НПФ в пространстве $\{\Phi_N, \Phi_D, \Phi_{ЕГЭ}\}$

Согласно параметрическому F-критерию кластеры групп ВУЗов различаются высоко значимо по Φ_N и $\Phi_{ЕГЭ}$ (уровень значимости $p < 0,0005$) и статистически значимо по Φ_D (на уровне значимости $0,005 < p \approx 0,018 < 0,05$). Непараметрический ранговый критерий Краскела-Уоллиса сглаживает различия до статистически значимых по Φ_N и $\Phi_{ЕГЭ}$ (уровень значимости $0,005 < p < 0,05$) и слабо незначимого по Φ_D (на уровне значимости $0,1 < p \approx 0,16$).

Согласно апостериорным параметрическим критериям можно выделить для каждого фактора однородные (различающиеся незначимо, то есть на уровне $p > 0,10$) группы кластеров, расположенные в порядке убывания факторных средних:

Φ_N : {K1}, {K2}, {K5, K4, K3}, {K4, K3, K6}, {K3, K6, K7}.

Φ_D : {K3}, {K4, K1, K6}, {K1, K6, K2, K5}, {K5, K7}.

$\Phi_{ЕГЭ}$: {K1, K6, K7}, {K7, K3}, {K3, K2}, {K4, K5}.

По непараметрическому ранговому критерию Краскела-Уоллиса:

Φ_N : {K1, K2}, {K5, K4, K3, K6, K7}.

Φ_D : {K3, K4, K1, K6, K2, K5, K7}.

$\Phi_{ЕГЭ}$: {K1, K6, K7, K3}, {K3, K2}, {K4, K5}.

Итоги кластерного анализа наблюдений по совокупности факторов, учитывая результаты множественных сравнений кластерных средних для каждого фактора, дают возможность провести классификацию наблюдений следующим образом: «Лидер» – (+2; +3), «Выше среднего» – (0,75; 2), «Средний» – (-0,75; +0,75), «Ниже среднего» – (-2; -0,75).

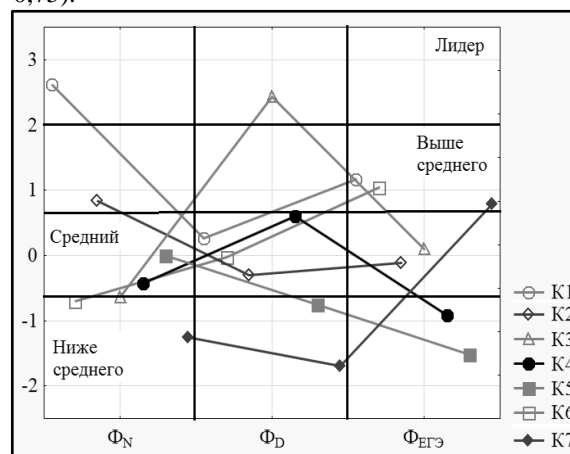


Рис.3. Линейные графики факторных средних
кластеров вузов по НПФ

Заключение

1. На основании корреляционного анализа выявлены 3 группы исходных показателей: $\{D_{ЕГЭ}\}$, $\{ЕГЭ_{14}, ЕГЭ_{15}\}$, и $\{N, N_C, N_{TS}, N_L, N_O\}$, что позволило с помощью факторного анализа построить 3-х факторную модель $\{\Phi_N, \Phi_D, \Phi_{ЕГЭ}\}$.

2. В рамках кластерного анализа в 3-х мерном факторном пространстве $\{\Phi_N, \Phi_D, \Phi_{ЕГЭ}\}$ построена 7-ми кластерная высококачественная модель 16-ти вузов по НПФ. Для каждого факторного показателя выделены группы однородных кластеров и оценен уровень значимости различий неоднородных кластеров.

3. Результаты кластерного анализа вузов по НПФ по совокупности показателей позволяют провести качественную классификацию вузов по НПФ в номинальной шкале измерений. Так, например, кластер K1 (МИФИ) является абсолютным лидером по числу обучающихся по НПФ; такие ВУЗы, как МГТУ, СПбПУ имеют уровень «выше среднего» по $\Phi_{ЕГЭ}$; а НИ ТПУ, входящий в кластер K2, имеет уровень «выше среднего» по Φ_N , «средний» - по Φ_D , «средний» - по $\Phi_{ЕГЭ}$.

Литература

1. Качество бюджетного приема в государственные вузы // URL: <http://www.hse.ru/ege/>

2. 2. Боровиков В.П. STATISTICA. Искусство анализа данных на компьютере: Для профессионалов. – СПб.: Питер, 2003. – 688 с.